PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/11, C07K 14/245, 14/36, 14/42, 14/79, C12N 9/06, 9/10, 9/90, G01N 33/68 — A1

(11) International Publication Number: WO 96/23879

(43) International Publication Date: 8 August 1996 (08.08.96)

(21) International Application Number: PCT/US96/01567

(22) International Filing Date: 30 January 1996 (30.01.96)

(30) Priority Data:
08/380,188 30 January 1995 (30.01.95) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/380,188 (CIP)
Filed on 30 January 1995 (30.01.95)

(71) Applicant *(for all designated States except US)*: TERRAPIN TECHNOLOGIES, INC. [US/US]; 750-H Gateway Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KAUVAR, Lawrence, M. [US/US]; 1205 Cole Street, San Francisco, CA 94117 (US). TRAINER, John, A. [US/US]; 4575 Niagara Avenue, San Diego, CA 92107 (US). VILLAR, Hugo, O. [US/US]; 7706 Hazelzut Drive, Newark, CA 94560 (US). NAPOLITANO, Eugene, W. [US/US]; 3575 Fillmore Street #303, San Francisco, CA 94123 (US).

(74) Agents: ADLER, Reid, G. et al.; Morrison & Foerster, 2000 Pennsylvania Avenue, N.W., Washington, DC 20006 (US).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GLUBODIES - MULTIPLICITIES OF PROTEINS CAPABLE OF BINDING A VARIETY OF SMALL MOLECULES

(57) Abstract

Certain naturally occurring proteins contain regions, designated protogludomains, which can be modified randomly or systematically to generate families of protein ligates, termed glubodies, which have ranges and variations in binding specificities comparable to those of the antibody repertoire. Methods for constructing such families of glubodies and for their use are described.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

GLUBODIES - MULTIPLICITIES OF PROTEINS CAPABLE OF BINDING A VARIETY OF SMALL MOLECULES

Technical Field

The invention relates generally to proteins that are capable of binding small molecules; and, more particularly, to the creation of families of proteins that result from randomization or other alteration of solvent-accessible loops substantially irrelevant to the remainder of the protein to confer on the family a range of binding affinities for small molecular targets.

Background

Natural selection in biological systems has resulted in the evolution of a number of macromolecules having the capacity to bind small molecular targets. Such macromolecules can be referred to as "ligates" in recognition of their ability to bind to cognate "ligands". Many naturally selected ligates, particularly protein ligates, exhibit very specific, high affinity binding with their cognate ligand. Typical examples include hormone receptors and their corresponding hormones; and enzymes and their corresponding substrates.

Such naturally occurring ligates, and/or their cognate ligands and analogs thereof, can be employed in a broad variety of applications; including analytical, diagnostic and therapeutic applications. However, the ligates available in nature for a particular purpose may have inappropriate specificities, be too costly to manufacture, or may have other physical properties that make them undesirable. Therefore, additional sources of ligates besides those that nature provides, would be desirable.

Various approaches to obtaining large families of additional potential ligates have been reported. Kaufman (International application PCT/CH85/00099) describes the generation of large numbers of proteins using random DNA sequences for recombinant production of these potential ligates. Ladner, U.S. patent 5,223,409, describes coupling of such variants to a genetically amplifiable unit, such as bacteriophage coat protein. Various alterations in antibodies to create subfamilies have also been attempted. The identification of oligonucleotides appropriate for use as

ligates is also described in PCT application WO94/08050 using selection procedures from large random mixtures of nucleic acids.

All of these approaches suffer from the exponential growth in numbers of possible variants as the number of monomers in the mixture of polymers increases -- i.e., a "combinatorial explosion."

Phage display libraries containing random mutations at positions 107, 108, 110, 111, 208, 213, 216, 219, 220 and 222 of GST A1-1 were prepared by Widersten, M., *et al.* J Mol Biol (1995) 250:115-122. Random mutations at positions 9-14, 102-112 and 210-220 of GST 2-2 were reported by Gorelick, A. *et al.* Proc Natl Acad Sci USA (1995) 92:8140-8144.

One approach to overcoming this combinatorial explosion has been described by one of the present inventors in U.S. patents 5,133,866 and 5,340,474. Systematic variation of the monomers results in a representative family using smaller numbers of polymers. Others have approached this problem by systematically varying, for example, one residue at a time. Huang, X. *et al.* Structural Biol (1994) 1:226-230 describe the preparation of a random library of myoglobin mutants prepared by using a single-base misincorporation random mutagenesis method. Palzkill, T. *et al.* in Proteins: Structure, Function and Genetics (1992) 14:29-44 describe a mutagenesis technique which randomizes the nucleotide sequence in a 3-6 codon region of a gene and then determines the percentage of random sequences that produce functional protein, where a low percentage of functionality indicates that mutageneic region is important for the structure and/or function of the protein.

Chimeric forms of glutathione transferases (GSTs) have also been prepared. Bjornestedt, R. *et al.* Biochem J (1992) 282:505-510 describe a human/rat chimera composed of a human alpha subunit I from the N-terminus to His143 or to Pro207 followed by the complementary C-terminal portion of rat alpha-I subunit. In addition, there have been studies to elucidate the function of particular residues in the isoenzyme GST P1-1. See Ricci, G. *et al.* J Biol Chem (1995) 270:1243-1248; LoBello, M. *et al.* ibid:1249-1253.

Nature's approach to generation of binding agents to a wide range of target small molecules is reflected in the generation of antibodies against virtually any target.

All vertebrates have large genomic loci for generation of the immunoglobulin repertoire wherein the variable regions of antibodies are assembled in response to stimulation by an antigen by rearrangement of individual portions of these loci to result in suitable binding characteristics. It is known that the regions specifically responsible for antigen binding, the complementarity determining regions (CDRs) are supported on a scaffolding of framework regions (FRs) and held in juxtaposition appropriate for antigen recognition. Immunoglobulins appear to be the only "family" of proteins that is known to have been created naturally to bind such a multiplicity of targets.

The present invention provides families of protein ligates, termed "glubodies", that are capable of binding a variety of small molecular ligands. Such families will be useful as sources of new ligates for the above described applications including analytical and diagnostic applications.

Disclosure of the Invention

The present invention provides families of potential ligates that are capable of binding a wide variety of small molecules. The families of the invention are obtained by taking advantage of a "loop" structure in a native protein and providing alterations in the loop to confer differing binding characteristics depending on the nature of the alterations.

The forms of the naturally occurring proteins that contain modified loops can be designated "glubodies", since they are, in a sense, analogs of antibodies which are capable of binding small molecules or what would correspond to a hapten. Since the glubodies are modified forms of naturally occurring proteins, these naturally occurring proteins can be called "protoglubodies". There are two regions of significance in the protoglubody -- a "protogludomain", which is the loop region, and the "framework" region. In the glubody, only the protogludomain has been modified.

Thus, in one aspect, the invention is directed to a method to prepare a multiplicity of member protein ligates (which collectively bind to or react with a variety of ligands), which method comprises identifying a protogludomain in a protoglubody protein and altering the protogludomain of each member of the family or multiplicity of protoglubody protein molecules. The alteration is different for each member. The

alterations in the protogludomain comprise substitution of one amino acid for another and/or deletion of one or more amino acids and/or insertion of one or more amino acids. Such alterations also include the randomized substitution of a segment of amino acids that are contiguous in the protogludomain, wherein said segment of amino acids comprises at least about three amino acids and fewer than about twenty amino acids, more preferably between about four and fifteen, still more preferably between about five and twelve. The randomized substitution is achieved by replacing amino acids with any member of a predetermined set of replacement residues.

Other aspects of the invention include the families of glubodies created by the method of the invention, families of nucleic acids encoding these families of glubodies, methods to produce the families by expressing the modified polynucleotides, and methods to utilize these families to select ligates or glubodies of particular desired properties, and in panels for analytical purposes.

Brief Description of the Drawings

Figure 1 shows the ribbon structure of human GST P-1-1, including S-hexyl glutathione docked at the binding site.

Figure 2 shows a pair-wise comparison of glubodies of the Gb/P204 family with native GST and with other members of the family.

Figure 3 shows the gray-scale representation data from Tables 2 and 3 (inhibition patterns of various Gb/P36 and Gb/P204 glubodies).

Figure 4 shows the gray-scale representation data from Table 4 (Gb/P204 and Gb/P206L glubodies).

Figure 5 shows the ribbon structure of retinol binding protein (RBP).

Figure 6 shows the ribbon structure of cyclophilin.

Modes of Carrying Out the Invention

The invention provides a means to obtain a multiplicity of "glubodies" which exhibit a range of binding properties analogous to the range obtainable in vertebrates by rearrangement of the immunoglobulin loci to obtain antibodies in response to a wide range of antigens. The invention takes advantage of the presence in various naturally

occurring proteins of "loop" structures, which are substantially irrelevant to the remainder of the protein and can be varied in amino acid sequence to generate a wide range of binding capabilities. The binding capabilities are not necessarily associated with the modified loop portion alone, but are related to the sequences occurring within that loop in relation to the remainder of the protein, which could be called a "framework".

The families of glubodies of the present invention are obtained by modification of the protogludomains of protoglubodies. A "protoglubody" refers to a protein having a "protogludomain" and a "framework." (More than one such protogludomain may be found in a single protoglubody.) A "protogludomain" is a region of a protein, containing 3-25 amino acids in a contiguous sequence, that: (i) is solvent-exposed or solvent-accessible; (ii) forms part of the cavity that defines a binding site; (iii) does not interact appreciably with residues outside the region; and (iv) in most cases, and preferably, lacks a well defined secondary structure.

Residues contained in a "solvent-accessible" domain are at least partly in contact with the bulk water surrounding the protein in solution. Solvent-accessibility can be assessed using any of a variety of techniques including, for example, the method of Conolly, M.L. Science (1983) 221:709. In many cases, solvent-accessibility will be apparent from visualization of the 3-D structure of a protein (i.e. in some cases a loop is apparent next to a concavity at the protein surface). Some proteins, such as HIV protease, are believed to have solvent accessible domains that are effectively opened and closed by a segment of the protein that functions as a "cover" (i.e. the cavity is covered by a segment that is capable of opening to allow entry of a ligand). Thus, these covered portions can also be considered solvent-accessible at the time the cover is open.

The location of a cavity that forms part of the binding site can be located using computer graphics such as those described by Levitt, D.G. *et al.* J Mol Graphics (1992) 10:229-234. This method displays protein cavities and their surrounding amino acids. Those cavities which are associated with binding sites can be identified by correlating the results of this method with standard direct techniques for locating binding sites of proteins.

A domain exhibits a lack of appreciable interaction with residues outside the domain when the residues in the domain have few if any structural contacts with other parts of the protein in which they reside (other than the peptide linkage). Thus, individual amino acid residues in such a topologically independent domain, exhibit few secondary structural contacts, such as H-bonds, X-X interactions or salt pairs, with amino acid residues in the framework outside of the protogludomain. While any limitation on the specific number of interactions would necessarily be arbitrary, it appears that less than three hydrogen bonds or salt bridges would be an acceptable cut-off point; preferably, no interactions of this type are exhibited.

The protogludomain also preferably exhibits little or no secondary structure. Lack of such structure is characteristic, for example, of β-turn regions and looser "C"-like structures.

Thus, the protogludomains of the invention represent structures that can loosely be referred to as "loops". The individual amino acids in a loop tend to exhibit fewer secondary structural contacts with neighboring amino acids (as compared to regions exhibiting pronounced secondary structure such as helices, sheets and hydrophobic cores). Preferred loops of the present invention tend to form part of a cavity and can be found on or near the surface of the protein. Preferred loops of the present invention also generally exhibit larger than average temperature factors (when the crystal structure is known). Correspondingly, in molecular dynamics simulations, preferred loops tend to exhibit larger than average atomic motions along a trajectory.

The "framework" refers to a portion of the protoglubody outside of the protogludomains that can act as a "scaffold" onto which a modified protogludomain -- i.e., a gludomain can be grafted using the techniques of the present invention. The primary function of the framework is to effectively display the protogludomain or gludomain on its surface so as to be available for binding target molecules.

Thus, a preferred framework for the generation of glubody families is contained in a protein that is known to:

1) bind peptides and/or small organic (haptenic) compounds;

2) have an active site/binding loop amenable to modification;

3) be monomeric with reasonably low molecular mass although it may be assembled into homo and heterodimers or multimers;

4) be stably expressed in the periplasm of *E. coli* or secreted if possible;

5) have a binding site distal to the N or C terminus; and

6) tolerate minor modifications of N- and C-termini, for example, for purification or detection.

The segment targeted -- the protogludomain -- 1) should be directly in or directly adjacent to the active site of the enzyme or protein; 2) should be directly involved in the binding of substrate/ligand molecules; 3) should not be part of the hydrophobic core of the protein and thus crucial to the structural integrity; 4) is preferably in a flexible loop, notably a β-turn region.

All of the foregoing properties may be established by examination of available crystallographic structural data, by computational chemistry or by homology modeling.

To summarize, a "glubody" refers to a modified form of a naturally occurring protoglubody protein having an unmodified framework and at least one modified protogludomain; whereby the resulting gludomain confers altered binding specificity of the glubody relative to its corresponding protoglubody. The glubody may be an assembly of monomers, one or more than one of which contains a modified protogludomain. Thus "glubody" may refer to a modified monomer or to such assemblies.

A "family" of glubodies refers to a multiplicity of different member glubodies derived from a single protoglubody, which differ from each other in the possession of different gludomains.

A "panel" of glubodies refers to a multiplicity of glubodies that may be, but need not be, members of a single family.

A "systematically-diversified panel" of glubodies refers to a panel whose individual members have been selected such that the panel members are collectively capable of binding to a wide variety of other molecules, but wherein there is a relatively low level of overlap in the binding specificity of particular panel members.

Another way to describe this maximal systematically arrived at diversity is in terms of the number of principal components needed to capture 50% of the variance in

binding of the glubody panel to a set of compounds; that set of compounds would include, for example, a set of compounds to which the parent protoglubody and any naturally occurring homologs bind. If the values of IC_{50}s for this set of compounds changes idiosyncratically for each of the glubodies in the panel, this increases the number of principal components. On the other hand, for a panel which contains a large number of members which do not react at all with the compounds, only one principal component accounts for more than 50% of the variance -- i.e., live or dead enzyme.

Identification and analysis of protoglubodies and protogludomains

A preferred source of protoglubodies includes proteins that are already known or believed to be ligates, i.e. proteins that are capable of binding other molecules. Among protein ligates, it is most convenient to focus on proteins for which structural data are available. Molecular modelling analyses in particular are helpful in initially assessing and provisionally confirming that a particular protein will be useful as a protoglubody. Protein structures can be assessed using any of a variety of techniques including, for example, X-ray crystallography, neutron diffraction and nuclear magnetic resonance. Especially preferred are high resolution crystallographic data based on crystallization of the protein in combination with a ligand bound to the protein; as exemplified in the cases of the human GST-P1-1, the rat GST 3:3, retinol binding protein, and cyclophilin discussed below. Analytical methods that provide guidance on the overall structure, even when it has not been solved experimentally, can also be employed. Such methods include, for example, homology modeling based on the solved structures of related or structurally similar proteins, as exemplified below.

With protein structure data available, the first step is the visual inspection of the three-dimensional structure to identify any protogludomains. For this purpose, the Insight II molecular modeling package available from Biosym Technologies, Inc. San Diego, CA is suitable.

A convenient initial screen can be performed by displaying the polypeptide framework of a prospective protoglubody without displaying the associated side chains. Suitable displays include framework "wire" models (in which lines connect all atoms along the polypeptide framework) and, more preferably, "ribbon" models (in

which the framework is drawn as a ribbon exhibiting turns and helices in the protein structure). In such displays, protogludomains tend to appear as relatively open domains, especially open loops, that are somewhat isolated from the remainder of the polypeptide framework.

Another useful and easily generated tool for identifying protogludomains includes maps illustrating pair-wise distances between all alpha carbons in a prospective protoglubody protein. In such maps, the alpha carbons of protogludomains tend to exhibit fewer near neighbors than the alpha carbons of other regions of the protein.

Preferred protgludomains can also be identified as regions exhibiting relatively large temperature factors as crystallographically determined. Moreover, using molecular dynamics simulations, preferred loops can be identified as regions exhibiting larger than average atomic motions. These motions can be determined using most commercially available molecular modeling software, including Discover, BiosymTech, Inc., San Diego, CA.

Regions that are part of the hydrophobic core of the protein, regions with extensive secondary structure, and regions with extensive secondary structural contacts with other parts of the protein are considered part of the framework as well as are polar residues that are buried (e.g. residues at the interface where oligomerization occurs).

In some instances, the protogludomain loop comprises a β-turn. Certain amino acids, particularly Gly, Ser and Tyr, have a tendency to appear in β-turn regions, see Chou P. Y. and Fasman, G. D., Annual Review of Biochemistry (1978) 47:251-276. As noted below, our analyses of the amino acid distributions in a wide variety of protein ligates suggest that these same amino acids (i.e. Gly, Ser and Tyr) are overrepresented in ligand binding sites.

The most preferred protogludomains of the present invention are solvent-accessible loops that are known to bind ligands and that appear to be quite independent topologically from the remainder of the protein.

In the absence of an experimentally determined three dimensional structure of the prospective protoglubody (or a similar protein), some relevant structural

information can still be obtained on the basis of the primary structure (i.e. the protein sequence). Thus, secondary structure prediction methods in combination with hydrophobicity plots can be used to identify structural motifs that are likely to be solvent exposed (see, e.g., the analytical methods described by Fasman, G.D. *et al.*, Trends Biochem Sci (1989) 14:295-299; and Benner, S.A. *et al.*, Curr Op Struct Biol (1992) 2:402-412). Such techniques can be used in conjunction with our sequence analytical techniques, described *infra*.

Differential distributions of particular amino acids provide additional information to predict the position of ligand-binding pockets. We analyzed the amino acid distribution patterns at ligand binding sites for 50 diverse protein ligates for which crystallographic data with bound ligands was available. We found Trp and His were present 250% and Arg and His 200% more frequently in contact with the ligand than their average observed across all proteins. More modest increases were observed for Ser and Asp. Conversely, Pro occurred much less frequently in proximity to the ligand binding site compared to its overall average. Other residues with decreased frequencies were Lys, Glu and Ala. Furthermore, Gly, Ser, Arg and Tyr were the most abundant residues within 4/ from the ligand. Thus, protogludomains tend to contain these amino acids.

In some cases, additional evidence will be available suggesting or confirming that a candidate protogludomain is involved in the binding of small molecular targets. For example, point mutation analyses may have identified residues likely to be involved in binding.

Thus, depending on what sort of information is available or readily obtainable for a prospective protoglubody, one or more of the aforementioned techniques can be employed to identify and analyze putative protogludomains within the protein. Illustrative examples of the use of these techniques in the context of several different types of proteins are described below.

A preferred subclass of protoglubodies will exhibit several other features that make subsequent steps in the manipulation of glubodies particularly convenient. Such preferred characteristics include: the ability to be stably expressed in *E. coli*; the ability to be transported to the bacterial periplasm for surface expression on filamentous

bacteriophage particles; the ability of the amino- or the carboxy-terminus to be modified with a tag for efficient purification (e.g. a hexahistidine tag); and a relatively small molecular size (preferably less than about 80 kD, more preferably in the 10-50 kD range).

Particularly preferred sources of protoglubodies are proteins that already bind to a number of small molecules. Among these are the so-called "protective" proteins that function in mammalian systems in the detoxification of exogenous substances and metabolic byproducts of oxidative metabolism.

Especially prominent among such enzymes are the glutathione S-transferases (GST's; Enzyme No. EC 2.5.1.18), GST's comprise a family of homodimeric cytosolic enzymes that catalyze the conjugation of glutathione (GSH) to a broad range of hydrophobic electrophiles. This reaction is one of the first steps in the inactivation and subsequent elimination of toxic xenobiotics which gain entry into the cell. The putative binding pocket of GST's has been proposed to consist of a GSH-binding domain, the G-site, and a second site, known as the H-site, believed to interact with hydrophobic compounds. A number of GSTs have been crystallized and characterized in the presence of an inhibitor that binds to the putative active site. The crystal structure of human GST-P1-1, a class Pi GST from human placenta, has been characterized in complex with S-hexyl glutathione at 2.8Å resolution, Reinemer, P., *et al.* J Mol Biol (1992) 227:214-226.

Modification of Protogludomains

After identifying a protogludomain that is to be targeted, the mechanical steps of mutagenesis can be performed using any of a variety of techniques. Most conveniently, however, PCR mutagenesis and related techniques can be used to randomize or otherwise alter residues in the region to be targeted. Such randomization or alteration can be readily tailored to result in the replacement of individual residues with any member of a pre-determined set of replacement residues. The replacement set can include, for example, the entire set of natural amino acids or pre-determined subsets thereof. Mutagenesis of the protogludomain can also include, for example, the addition or deletion of one or more residues within the domain. Such an approach can

be used to further expand the binding affinities of the families of resulting glubodies by generating family members having binding pockets of different sizes.

Although PCR-based mutagenesis is illustrated in the examples hereinbelow, alternative methods for synthesizing the glubodies of the invention can readily can be envisioned. Although most of them are substantially less convenient, they are at least theoretically possible. For example, the DNA encoding the entire glubody may be synthesized *de novo*, or only portions thereof may be synthesized *de novo* and ligated to portions obtained from cDNA or genomic DNA. The glubodies of the invention can be synthesized individually in this manner, or, as described below, an entire family are synthesized at once. One additional approach involves the use of codon amidites as described in copending U.S. application Serial No. 08/344,820 filed 23 November 1994, incorporated herein by reference. The use of protein synthesis techniques is also theoretically possible.

Assuming the most convenient method for preparing glubody families is applied -- namely, modification of the protogludomain by altering the amino acid sequence thereof at the DNA level and producing the resulting glubodies in recombinant host, individual colonies of host cells are cultured to obtain a library containing the members of the glubody family. The members of the glubody family can then be tested for ability to bind small molecule target candidates using standard biopanning or immunoassay-type techniques. Specific embodiments of these techniques are illustrated in the examples below. Depending on the manner in which glubody-encoding genes are expressed, the glubodies themselves may be displayed at the surface of the cells and/or phagemid particles secreted into the medium, or produced intracellularly. The methods for recovering the individual member glubodies will vary depending on the nature of the expression.

The families themselves may comprise members wherein the protoglutodomains have been completely randomized, resulting in large numbers of family members, or, preferably, the modifications in the protogludomains can be designed to confer maximal diversity in the family members. Techniques and considerations for designing such modified gludomains are based on the considerations set forth in U.S. Patents 4,963,263; 5,340,474 and 5,133,866, all of which are

incorporated herein by reference. Briefly, consideration is given to the properties of the amino acid sequence in the resulting gludomain in terms of maximizing diversity by supplying monomers that are maximally diverse in regard to at least two parameters that affect binding ability. In addition, advantage can be taken of the propensity of biding sites to contain preferred amino acid residues as described above.

Use of Glubody Families and Panels

The glubody families of the invention have a diversity of characteristics similar to that exhibited by the full basal repertoire of antibodies produced by vertebrate species. Accordingly, panels of glubodies can be used in a manner similar to antibodies in panels which are capable of fingerprinting individual compounds, matching patterns of fingerprints to determine binding capabilities of candidate compounds, and in identifying ligand-ligate pairs. These techniques are already described in detail in the art and need not be repeated here. Determination of molecular fingerprints for characterizing a single analyte and using these fingerprints to identify a candidate with qualities similar to those of known compounds is described in detail in U.S. Patents 5,217,869 and 5,300,425, both incorporated herein by reference.

As described, a single analyte can be characterized by obtaining a profile of reactivities of the analyte with the various glubody members of the panel. The characteristic pattern which emerges uniquely describes the analyte in question. The pattern of reactivities can either be determined by directly measuring the interaction of each member of the panel with the analyte, or by using a competitive technique described in the above-referenced patents. In the competitive technique, a diverse mixture of mimotopes, which mixture reacts essentially uniformly with each member of the panel is labeled and used to compete with the analyte to measure reactivity with respect to each member.

The use of such pattern matching techniques for analytical purposes is also described in U.S. Patent 5,338,659. Use of limited numbers of the glubody family in analytical techniques is also described in U.S. Patent 5,356,784. All of these patents are incorporated herein by reference.

In all of these applications, the panels may contain members of a single family of glubodies or can be comprised of members from two or more families. The selection of panel members depends on the application and availability.

Individual glubodies can also be used as members of reference panels used in technologies which translate binding capabilities of known compounds to screen small molecules for similar binding activities as described in U.S. Serial No. 08/177,673, filed 6 January 1994 and U.S. Serial No. 08/308,813, filed 19 September 1994. The disclosures of both applications are incorporated herein by reference.

In addition, individual glubodies can be used as affinity reagents, targeting agents, drug delivery vehicles, and the like, and in general in any manner that antibodies or immunologically reactive fragments of antibodies can be used. Since the glubodies also retain, in many instances, the ability to catalyze chemical reactions, they can be used as catalysts in a manner similar to that of their parent protoglubodies, with the added advantage that binding specificity and inhibition profiles can be altered in a manner appropriate for a particular set of reaction conditions.

Still another use for the glubodies of the invention is as general catalytic reagents. It is well known that antibodies can be used as catalysts in certain instances.

A summary of antibodies as catalysts can be found, for example, in an article by Lerner, R.A. *et al.* Science (1991) 252:659-667.

Glubodies are expected to be superior catalysts as compared to antibodies because the binding cleft of antibodies is relatively shallow as compared to that of most protogludomains, and indeed as compared to most enzymes. Therefore, in the glubodies of the invention, greater surface area is available for binding.

Particular antibodies have shown a modest ability to catalyze any of a wide range of reactions; the same range should be available with respect to glubodies. Thus, the nature of the catalytic activity is not limited by any original catalytic function of the protoglubody. Furthermore, the deeper clefts in glubodies should allow for better catalytic rate increases.

Finally, expression of genes encoding glubodies, especially intracellularly, provides a method for modulating the metabolism of the cell -- in essence as a gene

therapy tool. Others have provided precedents for this approach. For example, Carlson, J.R. in Mol Cell Biol (1988) 8:2638-2646 reports the intracellular expression of antibodies capable of neutralizing the endogenous yeast enzyme ADH1 *in vivo*. The cDNAs encoding the heavy and light chains were modified to remove signal sequences to prevent secretion, and the antibodies produced intracellularly were shown to neutralize the activity of the target enzyme *in vivo*.

Another report which demonstrates that intracellularly produced proteins can successfully interact is by Chien, C.T. *et al.* Proc Natl Acad Sci USA (1991) 88:9578-9582. Yeast cells were provided with an expression system containing a binding site for the GAL4 transcription activator protein upstream of a reporter protein-encoding sequence, so that expression of the reporter would occur only if the GAL4 protein were produced. The GAL4 protein has two regions, both necessary for its function -- a binding region and a polymerase activating region. These two regions were encoded on separate plasmids as part of coding sequences for chimeric proteins where the non-GAL4-related portion of the one of the chimeras was an amino acid sequence designed to interact with the non-GAL4 portion of the other chimera. Using the SIR4 protein which is known to form homodimers as the binding region of the chimeras, this system was successfully used to effect expression of the reporter gene, thus demonstrating that the two subunits of the homodimer interacted *in vivo*.

Thus, in a manner similar to that described above, expression systems for production of the glubodies of the invention can be used to modify cells so that the glubodies produced interact with targeted proteins that modulate metabolism. Such modulation could occur not only by binding targets intracellularly, but also by providing unique catalytic activity as described previously.

This last possibility also has some precedent. Gudkov, A.V. *et al.* Proc Natl Acad Sci USA (1993) 90:3231-3235 reported the isolation from cells resistant to drugs that act on topoisomerase-II (Topo II) a total of 12 different suppressor elements that corresponded to short segments of the Topo II α molecule or antisense RNA sequences which prevented Topo II gene expression.

Expression systems for the glubodies of the invention, therefore, in a similar manner can be used to alter the characteristics of a cell, including drug resistance.

EXAMPLES

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Sambrook, Fritsch, and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, Second Edition (1989), OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait Ed., 1984), ANIMAL CELL CULTURE (R.I. Freshney, Ed., 1987), the series METHODS IN ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J.M. Miller and M.P. Calos Eds. 1987), HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, (D.M. Weir and C.C. Blackwell, Eds.); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, (F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Siedman, J.A. Smith, and K. Struhl Eds. 1987); CURRENT PROTOCOLS IN IMMUNOLOGY (J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach and W. Strober Eds. 1991); and PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS (M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White Eds. 1990).

The examples presented below are provided as a further guide to the practitioner of ordinary skill in the art, and are not to be construed as limiting the invention in any way.

Example 1

Identification and Analysis of Two ProtoGludomains in a Human Glutathione S-Transferase (GST-P1-1) ProtoGlubody

In order to identify potential protogludomains in the human GST-P1-1 protein, we analyzed the protein for presence of the characteristics of gludomains as described above.

Using the Insight II molecular modelling package available from Biosym Technologies (San Diego, CA), we first visualized the three-dimensional structure of human GST-P1-1 using a "ribbon model" based on the crystal structure described by Reinemer, P., *et al.* J Mol Biol (1992) 227:214-226. It had been observed that the

folding topology, overall structure and subunit association of human GST-P1-1 closely resembled the structure of the porcine class Pi GST's.

The ribbon model of human GST-P1-1 is illustrated in Figure 1. As is apparent from the model, human GST-P1-1 contains an open loop (highlighted in Figure 1) located at the outer surface of the protein, that is relatively detached from the remainder of the protein framework. Moreover, from the data obtained by crystallizing GST-P1-1 in the presence of an inhibitor (S-hexylglutathione), Reinemer (*supra*), this loop is believed to be adjacent to the binding site for electrophilic substrates of the GST-P1-1 enzyme (the S-hexylglutathione molecule is also depicted in Figure 1).

A more detailed analysis of the putative protogludomain was conducted in order to identify particular residues that could be suitably targeted by loop mutagenesis. The entire segment from residues 36-43 appeared to form part of the loop. Analysis of temperature factors revealed that this region exhibited higher than average temperature factors indicating a high degree of flexibility, as expected for an exposed region with little secondary structure.

In the case of the protoglubody GST-P1-1, convenient and well-known assays for catalytic activity were available, and detailed information about the catalytic site was also available. We therefore decided, for this initial example, to further focus the modifications on residues that were likely to alter the binding of small xenobiotic compounds, without affecting the binding of the conjugation substrate glutathione. Thus, we focused on residues that were close to the S-hexyl group, but not likely to be directly involved with glutathione binding. Among the residues within 4Å of the S-hexyl group are: Tyr-7, which is likely to involved in catalysis; Tyr-106, which is in the region that has been implicated in the formation of salt bridges that stabilize the dimeric structure of the GSTs; and Val-35, which is part of an α-helical portion that is believed to form part of both the glutathione binding site and the putative xenobiotic binding site (β-site). Lys-44 is also believed to be important for glutathione binding as it is believed to form a salt bridge with the carboxylate terminus of glutathione.

The segment from Glu-36 to Leu-43 (the "36-43" loop, which has the sequence ETWQEGSL) was selected as the site to be altered. Although Trp-38 is believed to

act as a hydrogen bond donor to the Gly residue of glutathione, this single interaction should be a relatively weak one.

Further evidence suggesting that the region just beyond Val 35 might be useful as a gludomain came from homology studies and from information about the prospective binding sites in GST-P1-1. First, the most significant differences between the human and porcine Pi-class crystal structures are in this region, with the human form containing 2 residues that are lacking in the porcine form, a change that does not appear to prevent enzyme activity. Thus, individual residues in this region are not important to the overall function and stability of the protein. Second, while the GST isozymes from other classes resemble the Pi form in overall subunit folding topology and subunit association, the secondary structure of the region around Val 35 is markedly divergent, with no α-helical character at all for this segment in the published Mu-class isozyme Ji, X. *et al.* Biochemistry (1992) 31:10169-10181.

As described below, we created a glubody family (the "Gb/P36" family) by randomizing all of the amino acids in the segment from Glu-36 to Leu-43, using all of the natural amino acids as the set of "replacement residues". This loop mutagenesis, in which a very large variety of new loops are effectively grafted onto the protoglubody framework, was conveniently achieved using PCR mutagenesis, as described below.

Another protogludomain selected for modification was identified in the C-terminal region of GST-P1-1 as a loop comprising the residues from Ile-204 to Gln-210 (the "204-210" loop, which has the sequence INGNGKQ). The 204-210 loop is near the same cavity referred to above with respect to the 36-43 loop. Although the C-terminal 204-210 loop exhibited a higher than average temperature factor, it was not as high as that observed in the case of the 36-43 loop, suggesting that this C-terminal loop would not exhibit as much flexibility as the 36-43 loop.

A second glubody family, designated the Gb/P204 family, contained similar alterations in this loop. (We also created a glubody family (the "Gb/P36/P204" family) in which the glubodies contained modifications at both the 36-43 loop and the 204-210 loop.)

Finally, a glubody family was prepared wherein a randomized five amino acid sequence followed by a proline residue was inserted between residues 206 and 207. This family was designated Gb/P206L.

Example 2

Synthesis of the "Gb/P36" Family of Glubodies

The 36-43 loop in the human GST-P1-1 was completely randomized using PCR mutagenesis as described by Barbas *et al.* Proc Natl Acad Sci USA (1992) 89:4457-4461.

The complete sequence of human GST-P1-1 cDNA is available on GENBANK X06547. Human GST-P1-1 cDNA in the expression vector pKXHP1 was obtained from Kolm, R.H. *et al.* as described in "Gluthathione S-Transferases: Molecular Cloning, Site Directed Mutagenesis and Structure-Function Studies", G. Steinberg thesis, University of Stockholm (1991), and used as a PCR template. Since SfiI was to be used in subsequent cloning steps, as described below, we eliminated an internal *SfiI* site (at position 573-585 in the human P1 cDNA) using overlap PCR mutagenesis that employed the primers GST-P1SfiImutRC (see Table 1 for all primers). A G-to-A substitution at position 582 removed the *SfiI* site while leaving the amino acid sequence unchanged.

TABLE 1
Primers

Primer Name	Primer Sequence
GST-P1SfiImutRC	5'-TCAGGGGAGGCTAGGAGGCCTTGA-3'
GST-P1SfiImut	5'-TCAAGGCCTTCCTAGCCTCCCCTGA-3'
GST-P1SfiI	5'-CATGCCATGACTCGCGGCCCAGCCGGCCATGGCATGCCTCCATACACAGTTGTTTA-3'
GluPi-2	5'-CACGGTCACCACCTCCTCCTTCCA-3'
GluPi-1	5'-AAGGAGGAGGTGGTGACCGTGNNSNNSNNSNNSNNSNNSNNSAAAGCCTCCTGCCTATACGGG-3'
GST-P1MODNotI	5'-CCAGCATTCTGCGGCCGCCTGTTTCCCGTTGCCATTGATGG-3'
GST-P1MODNotIrandom	5'-CCAGCATTCTGCGGCCGCSNNSNNSNNSNNSNNSNNGGGGAGGTTCACGTACTCAGG-3'
GST-Loop-Pi	5'-CCAGCATTCTGCGGCCGCCTGTTTCCCGTTGCCGGGSNNSNNSNNSNNSNNATTGATGGGGGAGGTTCAC-3'

Two primary amplifications were performed. Reaction 1 contained 10 pmol each of primers GST-P1SfiI and GluPi-2, Perkin-Elmer *Taq* polymerase buffer (with 2 mM $MgCl_2$), 10 ng of template pKXHP1, all four dNTP's (250 µM each), and 2.5 units of *Taq* polymerase, in a final volume of 50 µl. Reaction 2 was identical to reaction 1 except that it contained the primers GluPi-1 and GST-P1MODNotI. Using an Omnigene thermal cycler, the reaction mixes were put through 25 cycles of denaturation (94°C, 1 min), annealing (65°C, 1 min), and extension (72°C, 1 min), followed by a final cycle of extension (72°C, 10 min). The reaction products were gel purified, subjected to overlap extension, and assembled as follows: 100 ng of purified

product from reaction 1 was combined with 100 ng of purified product from reaction 2, and then added to a PCR reaction mix containing *Taq* polymerase buffer (with 2 mM $MgCl_2$), all four dNTP's (250 µM each), and 2.5 units of *Taq* polymerase, in a final volume of 50 µl. This assembly mix was then taken through seven rounds of denaturation (94°C, 1 min) and annealing (65°C, 2.5 min), after which 10 pmols each of primers GST-P1SfiI and GST-P1MODNotI were added and the PCR amplification was continued for 25 cycles as above. The resulting product is DNA encoding a family of GST-P1 mutants with randomized loops in the position of the original 36-43 loop; designated "Gb/P36" cDNA. The cDNA fragment was gel purified, digested with *SfiI* and *NotI*, and gel purified once again.

The purified glubody cDNAs were ligated into a phagemid vector which can be used to facilitate expression of the glubodies either in the bacterial periplasm or on the surface of bacteria as fusions to the phage particle. The phagemid pHEN-1, described by Hoogenboom, *et al.* Nucleic Acids Res (1991) 19:4133-4137 was used for this purpose.

Digested cDNA (1 µg) was ligated (using a standard ligation reaction as described by Maniatis) to 1 µg SfiI/NotI-restricted pHEN-1. Ligated phagemid DNA was then electro-transformed into E. coli strain TG-1 by following established procedures Hoogenboom, *et al.* (*supra*). Transformants were spread onto two 150 mm 2X YT agar plates containing 100 µg/ml ampicillin for selection and 1% glucose, and incubated at 37°C overnight. Approximately 5 X 10^6 individual recombinant clones were generated. Cells were then scraped from the plates into 5 ml 2 X YT medium containing 100 µg/ml ampicillin and 1% glucose. 100 µl of scraped cells was used to inoculate 50 ml of 2 X YT including 100 µg/ml ampicillin and 1% glucose. This culture was grown at 37°C with shaking until the OD_{600} was approximately 1. It was then diluted 1:10 in the above medium and 5 µl of VCSM13 helper phage (7 X 10^{10} pfu/µl) added. The culture was incubated at 37°C for 15 min without shaking and further incubated at 37°C with shaking for an additional 45 min. Finally, this 50 ml culture was added to one liter of 2 X YT containing 100 µg/ml ampicillin and 50 µg/ml kanamycin and incubated overnight at 30°C.

The next day, phagemids were prepared by polyethylene glycol precipitation using the following protocol. The bacterial culture was then centrifuged at 4000 rpm for 10 min. at 4°C using a GSA rotor. The supernatant from this centrifugation was respun at 8000 rpm for 10 min. at 4°C in the GSA rotor. 0.15 volumes (150 ml) of 16.7% PEG/3.3M NaCl was added to this second supernatant. The solution was mixed well, placed at 4°C for one hour and spun at 8000 rpm for 30 min. in a GSA rotor maintained at 4°C. Phage pellets were resuspended in 40 ml dH_2O followed by the addition of 0.15 vol of PEG/NaCl. The solution was mixed well, placed at 4°C for 20 min. and spun at 8000 rpm in an SS34 rotor at 4°C. The supernatant from this spin was decanted and the phage pellet resuspended in 2 ml of sterile PBS. Resuspended phage were respun for 5 min. at 14,000 rpm in a microfuge and the supernatant filtered through a 0.45 um sterile filter. The phagemids constituting the library or glubody "family" were then titered and used in biopanning experiments described below.

Example 3

Glubody Purification Methods

To obtain purified glubody protein, the cDNA inserts of members of the family can be subcloned into pUC119Hismyc, a phagemid vector which directs the expression of cloned cDNAs as fusions to a six residue histidine tag which may be utilized in cheating affinity chromatography. Bacterial extracts are prepared as previously described, except that cells were resuspended and sonicated in column loading buffer (50 mM phosphate buffer pH 7.5; 500 mM NaCl; 20 mM imidazole). Extracts are spun at 10,000 rpm for 30 minutes at 4°C, and then supernatants are loaded onto a Ni-NTA resin column (Qiagen, Chatsworth, CA) and washed with 5 column volumes of 50 mM phosphate buffer pH 7.5, 500 mM NaCl, 35 mM imidazole. Glubodies are eluted with 50 mM phosphate buffer; 500 mM NaCl and 100 mM imidazole; and collected in six 1 ml fractions.

Randomly picked individual members of the glubody library, randomized in the region from residues 36 to 43, are screened for immunoreactivity in Western blot analysis to be sure the vector constructions had been effective. In most instances, recombinant proteins from bacterial extracts are detected with antibodies against both

GST-P1 and a fragment of *c-myc*, which links the C-terminus of glubodies with the N-terminus of phage gene III (data not shown).

Alternatively, glubodies can be prepared intracellularly and extracted. Bacterial cultures are centrifuged at 7000 X g in a Sorvall SS-34 rotor for 5 min. at 4°C. Cell pellets are frozen in dry ice/ethanol and resuspended in lysis buffer [10 mM Tris-HCl, pH 7.8, 50 mM EDTA, 15% glucose and 1 mg/ml lysozyme (Sigma)]. PMSF is added to a final concentration of 250 TM and the solution allowed to sit on ice for 1 hr. The suspension is sonicated for 2 min. with a Branson Sonifier 450 at 50% duty cycle and 6 output setting. Samples are centrifuged for 30 min. at 14,500 rpm (25,000 X g) in the Sorvall SS-34 rotor at 4°C. The supernatant is collected and stored at 4°C until further use.

Alternatively, an overnight culture of HB2151 grown in 2 X YT is diluted 1:50 and grown until the OD_{600} is 0.5. The culture was then infected with 2 µl of phagemid supernatant from recombinant glubodies originally propagated in bacterial strain TG-1. The culture was incubated for one hour at 37°C with shaking after which time ampicillin was added to final concentration of 100 µg/ml and the culture further incubated for one hour at 37°C with shaking. IPTG is then added to a final concentration of 1 mM and the culture incubated overnight at 30°C with shaking (225 rpm). Cultures are collected and cells pelleted at 14 K rpm in an Eppendorf microfuge. Supernatants are transferred to separate tubes for use in assays.

Example 4

Analysis of the Binding Properties of "Gb/P36" Glubodies

Gb/P36 glubodies were first tested for glutathione-S-transferase activity as measured by their ability to conjugate CDNB (1-chloro 2,4 dinitrobenzene) to GSH (glutathione), with an accompanying change in OD at 340 nm. Conjugation of GSH and CDNB was followed for 5 min at 30°C by measuring the absorbance at 340 nm in a thermostated microliter plate reader (Molecular Devices). Although the recombinant glubodies were quite similar by the structural analyses described above, they exhibited variation in their enzymatic properties. Of 30 randomly picked glubodies, 14 retained measurable enzyme activity in a control assay with the indicator substrate CDNB

(other clones were undetectable over background or had very low activities). Thus, approximately one half of the proteins in which the gludomain had been completely randomized retained the ability to bind CDNB and glutathione in their active sites, and were able to effectively catalyze the transferase reaction.

Further assessment of the binding properties of the novel glubodies was made by performing the catalytic reaction in the presence of sixteen potential inhibitors selected from different chemical classes. IC_{50}s were measured in the standard GST conjugation assay which contains 1 mM GSH and 1 mM CDNB in 200 mM sodium phosphate, pH 6.8. Compounds were assayed for their inhibitory activity at 250, 50, 10, 2 and 0.4 mM. The potency of the inhibitors used had previously been found to vary among the natural GSTs.

The IC_{50} data, summarized in Table 2 as -log IC_{50} (μM), demonstrate that a number of the resulting glubodies exhibited novel binding profiles with respect to this panel of compounds. (N.D. = not determined.)

TABLE 2
-log IC_{50} Values (TM)

Sequence Inhibitor	Native ETWQEG SL	Gb/36-1 FRRLLPS L	Gb/36-2 WRVVEIL V
Phloxine B	22.00	N.D.	N.D.
Fluoresceinamine, Isomer II	2500.00	2500.00	2500.00
Cibacron Brilliant Red 3BA	3.30	20.15	23.84
Fluorescein Isothiocyanate, Isomer I	146.00	613.17	340.89
9-Phenyl-2,3,7-Trihydroxy-6-Fluorone	2.80	463.46	2500.00
2-(4-(Fluorosulfonyl) Phenoxy) Acetic Acid	2500.00	2500.00	2500.00
Ibuprofen	2500.00	2500.00	2500.00
Cephaloglycin	73.36	112.16	162.01
hexyl-glutathione-Phenylglycine	0.97	2500.00	2430.64
Octyl Glutathione	7.10	201.44	204.85
(S)-6-Methoxy-A-Methyl-2-Naphthaleneacetic Acid	2500.00	2500.00	2500.00
1,2,3,4-Tetrafluoro-5,8-Dihydroxy-Anthraquinone	10.40	528.78	2500.00
Ranitidine	2500.00	2500.00	2500.00
6-Chloro-3-Nitro-2H-Chromene	0.10	0.44	0.17
Cholecalciferol	13.06	2500.00	2500.00
1,1'-Dibenzoylferrocene	10.22	N.D.	N.D.
1,1'-Dibromosalicil	0.51	116.65	121.67
Dienestrol	124.08	2500.00	2500.00

In order to demonstrate that activity in cell extracts was not influenced by contaminating bacterial proteins, one clone, Gb/204.3 (see below), was grown in larger quantities, purified by S-hexyl glutathione affinity chromatography, and retested. There were no significant changes in the binding profile of the partially-purified GP3

glubody preparation and the affinity-purified GP3, indicating that the binding properties of the crude extracts were indeed the result of GP3.

A more sophisticated analysis of the binding data in Table 2 provided further evidence of the significant changes in binding properties among the novel glubodies. Pair-wise comparison of the natural GSTs from different classes reveals significant overlap in their specificities, with correlation coefficients of 0.7 to 0.8. Comparing the novel glubodies described in Table 2 with the recombinant GST-P1-1 protoglubody, the highest correlation found was only 0.4. Moreover, particular glubodies such as Gb/36-1 and Gb/36-2 are also different from each other.

Example 5

Synthesis of the "Gb/P204" Library of Glubodies

For loop mutagenesis of the C-terminal 204-210 loop (INGNGKQ), a primary PCR reaction was performed as described above except that 10 pmols of the primer GST-P1MODNotrandom was used to replace GST-P1MODNotI in the amplification. PCR products were purified and digested as above. The mutant cDNA generated from this reaction was designated "Gb/P204" cDNA.

The members of this glubody family were produced in *E. coli* in a manner analogous to that set forth for the Gb/P36 family as described in Examples 2 and 3.

The glubodies of the Gb/P204 family were tested for gluthathione S-transferase activity as measured by ability to couple CDNB as described in Example 4. This catalytic reaction is performed in the presence of 18 potential inhibitors selected from different chemical classes to obtain IC_{50} data for these proteins. Again, approximately half of the glubodies retained this ability.

Illustrative members of this family and their IC_{50} values (-log IC_{50} (μM)) with respect to various inhibitors are listed in Table 3. (N.D. = not determined.)

TABLE 3
-log IC_{50} Values (μM)

Sequence Inhibitor	Native INGNG KQ	Gb/204-1 PEQHAP E	Gb/204-2 HPDPPQ A	Gb/204-3 MATGN R	Gb/204-5 GERRLE
Phloxine B	9.27	154.41	132.72	N.D.	103.76
Fluoresceinamine, Isomer II	2415.24	N.D.	N.D.	2500.00	N.D.
Cibacron Brilliant Red 3BA	1.10	59.40	119.79	4.11	37.54
Fluorescein Isothiocyanate, Isomer I	2500.00	352.88	314.00	2500.00	498.21
9-Phenyl-2,3,7-Trihydroxy-6-Fluorone	37.97	564.67	2500.00	2500.00	2500.00
2-(4-(Fluorosulfonyl) Phenoxy) Acetic Acid	2500.00	897.83	2500.00	2500.00	167.42
Ibuprofen	2500.00	2500.00	2500.00	2500.00	2500.00
Cephaloglycin	1527.76	665.02	1052.42	2500.00	428.30
hexyl-glutathione-Phenylglycine	3.57	331.60	N.D.	24.02	319.84
Octyl Glutathione	13.89	2500.00	2500.00	63.49	352.73
(S)-6-Methoxy-A-Methyl-2-Naphthaleneacetic Acid	2500.00	2500.00	2500.00	2500.00	2500.00
1,2,3,4-Tetrafluoro-5,8-Dihydroxy-Anthraquinone	61.35	113.18	196.56	2500.00	2500.00
Ranitidine	2500.00	2222.48	1758.29	2500.00	2122.09
6-Chloro-3-Nitro-2H-Chromene	1.18	2500.00	2500.00	6.84	2500.00
Cholecalciferol	2500.00	2500.00	2500.00	2500.00	2500.00
1,1'-Dibenzoylferrocene	69.92	257.93	1136.10	2500.00	227.23
1,1'-Dibromosalicil	12.66	N.D.	N.D.	N.D.	N.D.
Dienestrol	785.26	2500.00	2500.00	151.07	2500.00

As was done with respect to the Gb/P36 family, pair-wise comparison was made with respect to the native GST and between individual members of the glubody family Gb/P204. The results of these pair-wise correlations are shown in Figure 2. As shown, the correlation in binding properties among the various glubodies and the protoglubody varies from poor to nonsignificant. Similarly, the correlation among the glubodies themselves varies substantially. In contrast, natural isozymes of this protoglubody are known to be strongly correlated in binding properties despite primary amino acid sequence differences far larger than among the glubodies. This indicates that the protogludomain has a strong influence on binding properties.

Figure 3 summarizes the data of Tables 2 and 3 in "gray scale" form. In this scale, black boxes represent the most potent compounds and white boxes represent no detectable inhibition. P1-1 is the parental recombinant enzyme. recP1 is the recombinant enzyme expressed with a C-terminal *c-myc* tag.

Example 6

Synthesis of the "Gb/P206 Loop" Library of Glubodies

Another glubody library, the "Gb/P206L" library, was prepared by loop mutagenesis of the human GST-P1-1 protoglubody, to generate insertions expanding the size of the 204-210 loop. In this "expansion" loop mutagenesis, each glubody receives a novel loop comprising a single hexapeptide insertion between residues 205 and 206. The insertion contained five random amino acid residues followed by a proline residue. The resulting loops thus comprise the sequence IN(XXXXXP)GNGKQ. The cDNA for this library was generated under the same PCR amplification conditions described above except that the 3' oligonucleotide primer was GST-Loop-Pi (see primers in Table 1).

Table 4 provides IC_{50} values (-log IC_{50} (μM) for two members of the Gb/P206L family (Gb/8 and Gb/12) and three additional members of the Gb/P204 family (Gb/19, Gb/21, and Gb/23). Gb/19 showed a frame shift subsequent to residue 204; therefore all 19 amino acids downstream of this residue (as opposed to 6) were different from the native sequence. As might have been expected, this glubody shows unresponsiveness to most inhibitors, and is, thus, a perhaps unintentional control.

The values in Table 4 represent the mean ±1S.D. of three separate assays. N.I. = not inhibitory.

Table 4a shows correlation coefficients with respect to the glubodies assayed in Table 4. Again, Gb/19 shows comparatively low values whereas the remaining glubodies behave analogously to the native enzyme (P1-1) and to the recombinantly prepared dimer (rP1-1).

Figure 4 shows the results obtained in Table 4 as the "gray scale" described above.

Table 4 - Enzyme Inhibition (IC_{50})

#	RP	Compound Name	P1-1	recomb P1	Gb/8	Gb/12	Gb/19	Gb/21	Gb/23
1	16	Fluorescein isothiocyanate, isomer 1	360±194	N.I.	N.I.	214±33	N.I.	N.I.	N.I.
2	18	9-phenyl-2,3,7-trihydroxy-6-fluorone	11.5±3.4	79.1±27.0	166±56	94.1±40.2	N.I.	137±33	N.I.
3	24	4-(6-hydroxy-3-oxo-3H-xanthen-9-yl)benzoic acid	426±158	N.I.	N.I.	N.I.	337±197	N.I.	314±119
4	26	6-hydroxy-3-oxo-3H-xanthene-9-propionic acid	43.1±4.8	116±49	N.I.	412±227	257±138	137±51	45.8±4.9
5	134	1,2,3,4-tetrafluoro-5,8-dihydroxy-anthraquinone, 99%	16.3±0.7	294±167	77.9±4.4	7.8±1.0	N.I.	7.9±2.2	29.4±18.1
6	88	Cephaloglycin	17.5±9.9	N.I.	N.I.	N.I.	N.I.	398±71	301±32
7	91	Cephapirin sodium salt	41.7±9.1	123±29	N.I.	308±100	N.I.	86.9±6.2	N.I.
8	92	Cephalothin sodium salt	20.0±5.5	72.7±20.8	486±253	N.I.	474±125	44.4±8.3	213±47
9	95	Cephaloridine	27.4±10.3	77.0±11.1	389±168	183±62	481±136	42.4±5.8	343±134
10	4142	Cefoperazone sodium salt, 90%	39.4±1.1	153±52	328±12	258±99	N.I.	73.3±14.7	328±73
11	102	gE-C(Hx)-G	48.61*	310±147	N.I.	241±35	N.I.	193±37	115±54
12	107	S-2-methylnaphthyl glutathione	10.7±1.6	50.5±10.8	219±101	127±48	N.I.	136±31	67.1±8.9
13	4106	gE-C(Bu)-G	79.8±11.2	309±107	356±61	113±8	N.I.	287±74	196±107
14	4298	gE-C(Bz)-PG	3.0±3.2	23.0±7.7	140±29	141±50	N.I.	296±212	150±74
15	46	2-mandelamido-4-nitrophenol	225±125	N.I.	N.I.	N.I.	N.I.	405±101	308±111
16	1720	3-nitro-5,6-benzo-2H-chromene	<1μM	12.1±7.0	17.4±3.4	20.2±3.2	59.4±17.2	10.9±2.1	75.1±3.6
17	2655	1-benzene sulfonamideoxime-naphthaquim-4-one	19.6±2.0	101±64	N.I.	31.0±22.9	N.I.	78.8±22.1	84.1±8.8
18	3038	3-cyano-5-pentyl-2,3-dihydro-pyridine-2-thione	N.I.	N.I.	139±51	N.I.	N.I.	N.I.	N.I.

Table 4a - Correlation Coefficients

	P1-1	r-P1-1	Gb-8	Gb-12	Gb-19	Gb-21	Gb-23
P11	1.000						
rP11	0.836	1.000					
Gb8	0.470	0.541	1.000				
Gb12	0.592	0.591	0.502	1.000			
Gb19	0.291	0.356	0.254	0.025	1.000		
Gb21	0.749	0.720	0.478	0.647	0.307	1.000	
Gb23	0.503	0.390	0.175	0.412	0.314	0.547	1.000

Example 7

Synthesis of the "Gb/P36/204" Library of Glubodies

Using essentially the same techniques as described above, we created a glubody family, designated the "Gb/P36/204" library or family wherein loop mutagenesis at both the 36-43 loop was conducted as described in Example 2 in synthesizing the Gb/P36 library and the 204-210 loop as described in Example 5 to prepare the Gb/P204 family.

Example 8

Identification and Isolation of Glubodies Binding Particular Targets

Recombinant glubodies expressed on the surface of phagemid particles, as described above, are panned against desired ligands in a procedure essentially described by Marks, J.D. *et al.* J Mol Biol (1991) 222:581-597.

Briefly, 96-well plates are coated with 5 µg streptavidin per well in 100 µl coating buffer (0.1 M $NaHCO_3$) overnight at 4°C. The streptavidin solution is removed and replaced with blocking buffer (2% dry milk in PBS) for 30 min. at RT. Wells are then washed 5 X PBS/Tween (0.02%) followed by two washes with PBS and overlayed with 1 µg biotinylated candidate targets in 100 µl PBS for one hour at RT. Wells are washed as above and 10^{11}-10^{12} phagemid particles from rescued libraries added to each well followed by incubation for two hours at RT. Unbound

phage are removed and the wells washed 10 X PBS/Tween (0.02%) followed by two washes with PBS.

Alternatively, 10^{12} phagemids are preincubated with 1 ng - 1 µg biotinylated candidate target in solution for two hours at RT and then added to the streptavidin-coated wells for an additional 30 min. and washed as above.

Phagemids are then eluted with 100 µl 0.1 M HCl pH 2.2 containing 1 mg/ml BSA for 10 min. at RT. Samples are immediately neutralized with 7 µl 2 M Tris base. Eluted phagemids are amplified by infecting 2 ml TG-1 (OD_{600} = 0.8-1.0) with the entire elution and allowing the culture to sit at RT for 15 min. 10 ml of prewarmed (37°C) 2 X YT containing 40 µg/ml ampicillin is added and the culture grown at 37°C for one hour with shaking. The ampicillin concentration is adjusted to 100 µg/ml and the culture is incubated for 45 min. at 37°C in the shaker (250 rpm). Shaking is then slowed to 100 rpm for 15 min. to allow pili to regenerate. VCSM13 helper phage (10^{12} pfu) is added and the culture is incubated at 37°C for 15 min. without shaking. The culture was transferred to 200 ml 2 X YT containing ampicillin at 100 µg/ml and incubated in the shaker for one hour at 37°C. Finally, kanamycin was added to a final concentration of 50 µg/ml and bacteria were grown overnight at 30°C, 225 rpm.

Binding to a target candidate can also be tested using ELISA assays. 96-well plates are coated with 5 µg/well streptavidin in 100 µl 0.1 M $NaHCO_3$ pH 9.2 in a humidified chamber overnight at 4°C . The streptavidin solution is removed and replaced with 100 µl blocking solution (1%BSA/PBS) and incubated for 30 min. at RT. Wells are washed 5 X PBS/Tween (0.02%) followed by two rinses with PBS. One µg of biotinylated target candidate is added to each well in 100 µl PBS for one hour at RT. Wells are washed as above and 10^{10} phagemid particles from individual recombinants, polyclonal amplified phagemid populations or soluble glubody protein extract is added to the wells. Wells are washed as above and 100 µl of anti-M13 polyclonal IgG at a 1:1000 dilution or mAb 9E10 (Santa Cruz Biotechnology, Inc., Santa Cruz, CA; anti-c-myc tag) at 1 µg/ml in 2% dry milk/PBS was added for one hour at RT.

Plates are washed as above and secondary antibody *e.g* alkaline phosphatase-conjugated goat anti-rabbit or goat anti-mouse antibody diluted 1:1000 in

2%Milk/PBS] is added for one hour at RT. Wells are washed 3 X PBS/Tween (0.02%) followed by two rinses with dH_2O. The wells are then developed with 100 µl of 10 mM diethanolamine, 1 mM $MgCl_2$ containing 1 mg/ml pNPP (p-nitrophenyl phosphate) and read at 405 nm in an ELISA plate reader (Molecular Devices, Palo Alto, CA).

Example 9

Additional Protoglubodies

Structures having the features of protogludomains (as described above) are not very common, but they are relatively easy to detect using the methods disclosed here in proteins of a wide variety of different types, including proteins that would not otherwise be expected to be useful for these applications.

Further analysis of potential candidates using molecular modelling resulted in the elucidation of a number of different proteins likely to be useful for the production of glubodies using the techniques described herein. Several illustrations of these are described below.

A. Identification and Analysis of Gludomains in a Rat GST (GST 3:3) ProtoGlubody

The crystal structure of rat glutathione S-transferase ("GST 3:3") has been determined in a complex with glutathione at 2.2Å resolution. See Ji, X, *et al.* Biochemistry (1992) 31:10169-10181. The rat GST 3:3 data was obtained from the prerelease directory of the Brookhaven Protein Databank (Brookhaven code: 1GST).

Using the techniques described above, we were able to identify a potential gludomain at residues 32-48.

B. Human DHFR

Human dihydrofolate reductase (DHFR) exhibited at least two regions that are likely protogludomains: Lys-18 to Leu-22 (the "18-22 loop"), and Phe-58 to Arg-65 (the "58-65 loop"). In particular, both of these domains exist as solvent-exposed loops, and the limited hydrogen bonding engaged in by amino acids in these domains is

essentially restricted to amino acids within the loop. Thus, these solvent-exposed loops also exhibit topological independence.

In contrast, amino acids immediately outside each of the loops engage in more substantial interactions with other parts of the protein. Outside of the 18-22 loop, for example, Gly-17 forms a hydrogen bond with Asp-145, and Pro-23 interacts with a water molecule that in turn interacts with Ser-144.

Additional evidence suggesting that these two loops will be useful as protogludomains comes from crystallization studies. In particular, both the 18-22 loop and the 58-65 loop are situated within 10 Å from the bound folate molecule in the crystal structures studied determined by Davies II, J.F. *et al.* Biochemistry (1990), 29:9467 and are thus contiguous with a cavity in the protein that could act as a binding site.

Other features also contribute to making human DHFR a preferred protoglubody: the cDNA of human DHFR is available (Nienhuis, A.W. *et al.*, J Biol Chem (1984) 259:3933-43; the human recombinant protein has been expressed in E. coli; and the protein is of a relatively convenient size (it is a dimer comprising 186 residues per monomer).

C. Human Retinol Binding Protein

Retinol binding protein (RBP) is synthesized in hepatocytes, loaded with retinol, and secreted into the plasma where it serves to deliver retinol to various tissues and organs. Human RBP is a single-chain 21 kD protein and is a member of the lipocalin superfamily. Lipocalins are involved in ligand transport and include, besides RBP, β-lactoglobulin and bilin binding protein. Human RBP has been produced recombinantly in the cytosol (Wang, *et al.* Gene (1993) 133:291-294); in the periplasm of *E. coli* (Sivaprasadarao, *et al.* Biochem J (1993) 296:209-215); and in secreted form from *E. coli* (our unpublished results). The secreted form has been shown to retain retinol binding activity.

Human RBP has been characterized crystallographically by Cowan, S. *et al.* Proteins (1990) 8:44-61 and a ribbon structure for this protein is shown in Figure 5.

The structure shows two candidate protogludomains in the region Val61-Val69 and Gly92-Gln98. Both are solvent-exposed loops that form part of the binding site.

D. *E. coli* Biotin Repressor

Another illustrative example of a protein exhibiting protogludomains is the biotin repressor protein from *E. coli*. In particular, the region from Tyr-111 to Arg-118 is a solvent-exposed loop in the form of a β-strand segment that engages in few if any secondary structural contacts with other parts of the structure. The 111-118 loop lies on top of the biotin molecule.

The cDNA of the E. Coli biotin repressor is available (Otsuka, *et al.*, Gene (1985) 35:321-331 (1985); the protein has 321 residues; and the crystal structure has been solved (Wilson, K.P. *et al.* Proc Natl Acad Sci (1992) 89:9257).

E. Streptomyces Streptavidin

Another illustrative example of a protein exhibiting protogludomains is the streptavidin protein from Streptomyces. The region from Gly-113 to Lys-121 is a solvent-exposed loop that engages in few if any secondary structural contacts with other parts of the structure. The 113-121 loop lies near the entrance of the cavity that binds biotin. The residues in the loop are approximately 12 Å from the bound biotin molecule. While this distance is larger than in the examples above, the loop does form part of the cavity where the biotin molecule is bound, and alteration of the loop can be expected to result in changing the electrostatic properties within the cavity itself, which would be predicted to alter the binding profiles for the resulting glubodies.

The cDNA of streptavidin derived from Streptomyces avidinii is available (Cantor, *et al.*, Nucleic Acids Res (1986) 14:1871-1882; the protein comprises 159 residues; and the crystal structure is available (Weber, P.C. *et al.* Science (1989) 243:85).

F. Human Cyclophilin

The cyclophilins are a family of highly conserved proteins that display high affinity and binding to cyclosporin A, catalyze the cis/trans isomerization of a peptide

bond between proline and its N-terminal neighbor, and are thought to be involved in the late stages of protein folding. Human cyclophilin has been produced recombinantly in the cytosol of *E. coli* (Liu *et al.* Proc Natl Acad Sci USA (1990) 87:2304-2308 and a naturally occurring periplasmic *E. coli* cyclophilin homolog has been isolated (Liu *et al. ibid.* 4028-2032).

Human cyclophilin binds to peptides smaller than cyclosporin and has been crystallized in a complex with a tetrapeptide (Kallen *et al.* Nature (1991) 353:276-279). The crystallographic structure has been determined by Ke, H. *et al.* Proc Natl Acad Sci USA (1991) 88:7483. The region spanning Lys118-Lys125 has the characteristics of a protogludomain in that it is solvent exposed, forms limited interactions with noncontiguous residues and is part of the ligand binding site.

The ribbon structure of the rat counterpart of human cyclophilin is shown in Figure 6. Rat and human isozymes have 96% conserved amino acid sequence, and the rat cyclophilin cDNA is therefore used to create a glubody family in the Lys118-Lys125 region in a manner analogous to that described above for the creation of Gb/P36.

Claims

1. A method to prepare a family of member protein glubodies wherein said family binds to or reacts with a variety of ligands, which method comprises

identifying a proto-gludomain in a proto-glubody protein; and

effecting in said proto-gludomain of each member of a multiplicity of molecules of said proto-glubody protein, an alteration of the amino acid sequence, wherein said alteration is different for each member.

2. The method of claim 1 wherein said alteration includes substitution of one amino acid for another and/or deletion of one or more amino acids and/or insertion of one or more amino acids.

3. The method of claim 1 wherein said alteration comprises substitutions of 1-6 amino acid positions in said proto-gludomain randomized among said members; or

wherein said alteration comprises substitutions of 1-6 amino acid positions in said proto-gludomain designed to produce a maximally diverse multiplicity of glubodies.

4. The method of claim 1 wherein said alteration is effected by mutagenizing the nucleotide sequence encoding said proto-gludomain in each member of said multiplicity of molecules of said proto-glubody protein.

5. A method to obtain a protein ligate glubody of desired binding properties, which method comprises

identifying a proto-gludomain in a proto-glubody protein;

effecting in a proto-gludomain of each member of a multiplicity of molecules of said proto-glubody protein, an alteration of amino acid sequence wherein said alteration is different for each member; and

selecting from said multiplicity a protein ligate glubody of desired properties.

6. A multiplicity of protein glubodies, wherein said multiplicity binds to or reacts with a variety of ligands,

wherein each said glubody has a modification in the protogludomain of a protoglubody, and wherein each member of said multiplicity has a different altered amino acid sequence in its proto-gludomain.

7. The multiplicity of claim 6 wherein said altered amino acid sequence comprises substitution of one amino acid for another and/or deletion of one or more amino acids and/or insertion of one or more amino acids.

8. The multiplicity of claim 6 wherein said altered amino acid sequence comprises substitutions of 1-6 amino acid positions in said proto-gludomain randomized among said members.

9. The multiplicity of claim 6 which comprises a systematically diversified panel of glubodies.

10. The multiplicity of claim 6 which is a family of glubodies wherein said altered amino acid sequence is prepared by a process that comprises mutagenizing the nucleotide sequence encoding said proto-gludomain in each member of said multiplicity of molecules of a single protoglubody protein.

11. The multiplicity of claim 10 wherein said protoglubody is selected from the group consisting of a glutathione S-transferase, a retinol binding protein, a cyclophilin, a dihydrofolate reductase, a ferredoxin, a biotin repressor, a streptavidin protein and a ricin protein.

12. The multiplicity of claim 11 wherein said protoglubody is a glutathione-S-transferase or a retinol binding protein.

13. A composition of DNA molecules which encodes the multiplicity of claim 6.

14. A composition of DNA molecules which encodes the multiplicity of claim 10.

15. A composition of DNA which comprises expression systems effective in producing the multiplicity of claim 6.

16. A composition of DNA which comprises expression systems effective in producing the multiplicity of claim 10.

17. A method to characterize a single analyte, which method comprises:

contacting said analyte with each member of a panel of glubodies said glubodies having characteristics of the multiplicity of claim 6;

detecting the degree of reactivity of said analyte to each of said glubodies;

recording said degree of reactivity of said analyte to each of said glubodies; and

arranging said recorded degrees of reactivity so as to provide a characteristic profile of said analyte.

18. The method of claim 17, wherein said detecting is by reacting unlabelled analyte competitively with a diverse mixture of labeled mimotopes with respect to each of said glubodies, which mixture is approximately equally reactive with each glubody in said panel and measuring the reduction in binding of the labeled mixture to each glubody in the panel.

19. The method of claim 18 wherein said glubodies are coupled to a solid support in a predetermined pattern.

20. A method to identify a candidate, which candidate will be effective in reacting with a target, wherein said target has a known ligand with which it reacts, which method comprises:

contacting said candidate with each of a panel of glubodies, which glubodies react in a multiplicity of differing degrees with said candidate;

detecting the degree of reactivity of said candidate to each of said glubodies;

recording each said degree of reactivity of said candidate to each of said glubodies;

arranging said recorded degrees of reactivity so as to provide a characteristic profile of said candidate;

comparing said profile to a profile analogously obtained of said ligand with respect to said multiplicity of glubodies;

wherein similarity of the profile of said candidate to the profile of said ligand indicates the ability of the candidate to react with said target.

21. The method of claim 20 wherein said target is a receptor and the candidate is a candidate drug.

22. The method of claim 20 wherein said glubodies are coupled to a solid support in a predetermined pattern.

23. A method to identify a candidate reactive with a target, which method comprises:

(a) providing a formula that represents a combination of the reactivity profiles with respect to a first set of candidates of at least two members of a panel comprising the multiplicity of glubodies of claim 6, which formula calculates a predicted profile that best matches the reactivity profile of the target with respect to said first set of candidates;

(b) testing the reactivity of said at least two members of the panel with respect to a candidate; and

(c) calculating a predicted reactivity with respect to the target for said candidate by applying said formula to the reactivities determined in step (b) to estimate the reactivity of the candidate with respect to the target; and identifying a substance as being a candidate predicted to react with the target

24. The method of claim 23 which further includes the step of assembling the identified substance from starting materials appropriate to said substance.

25. A method to modulate the metabolism of a cell which method comprises culturing said cell, which has been modified to contain an expression system effective in producing a glubody identified by the method of claim 5 under conditions wherein said glubody is produced intracellularly so as to effect an interaction between said glubody and intracellular components to modulate said metabolism.

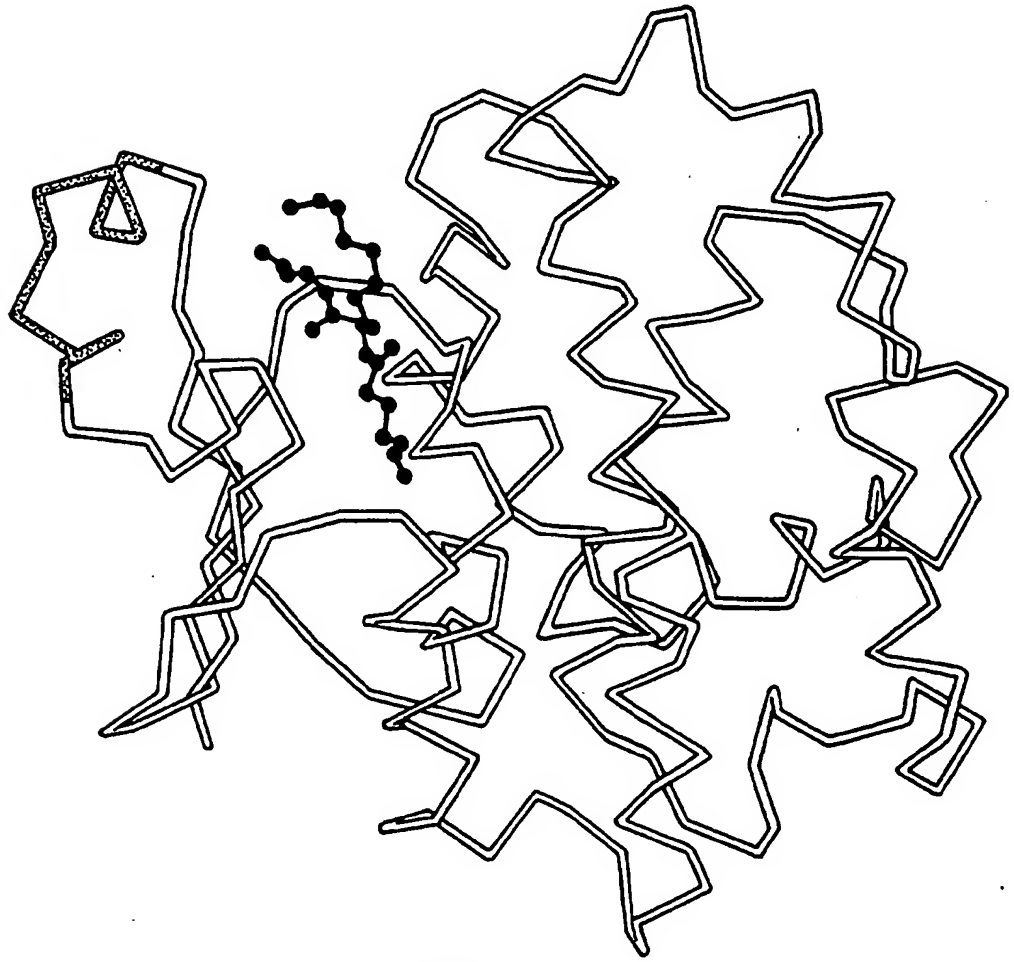


FIG. 1

FIG. 2A

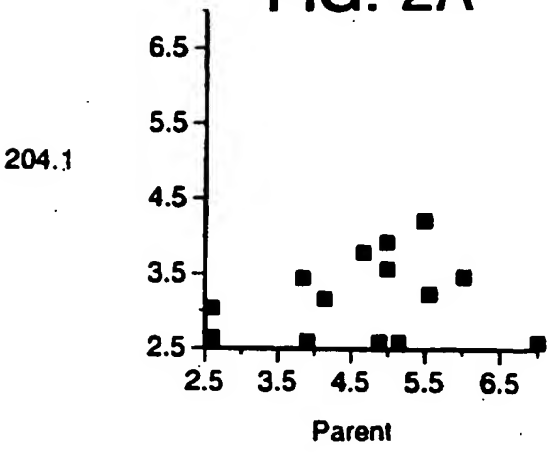


FIG. 2B

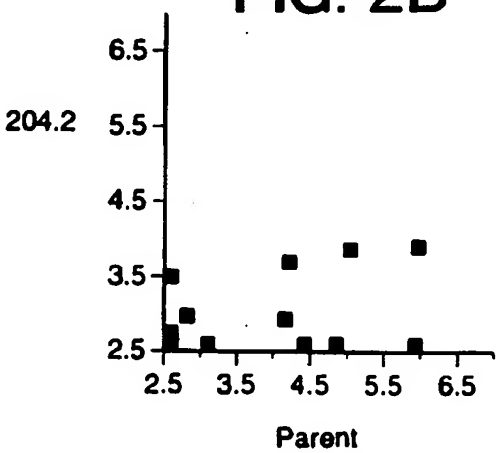


FIG. 2C

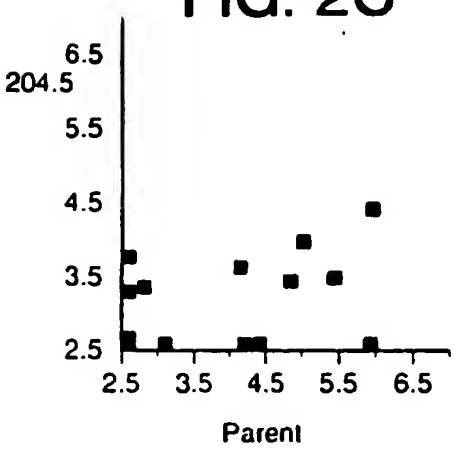


FIG. 2D

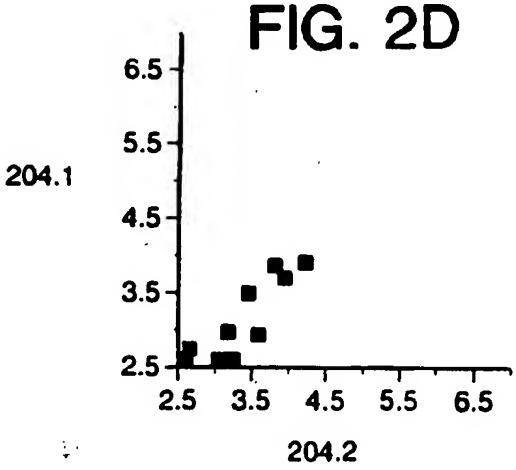


FIG. 2E

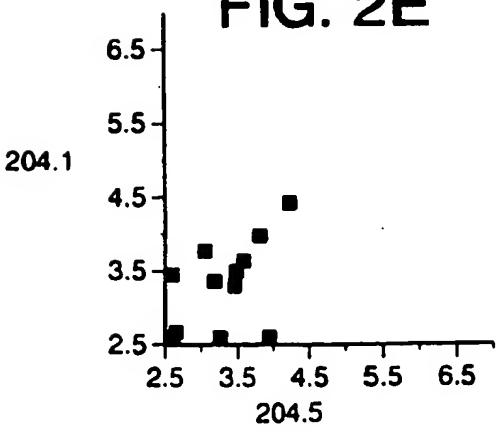
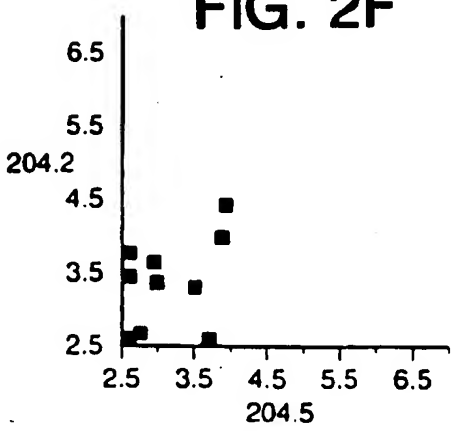


FIG. 2F



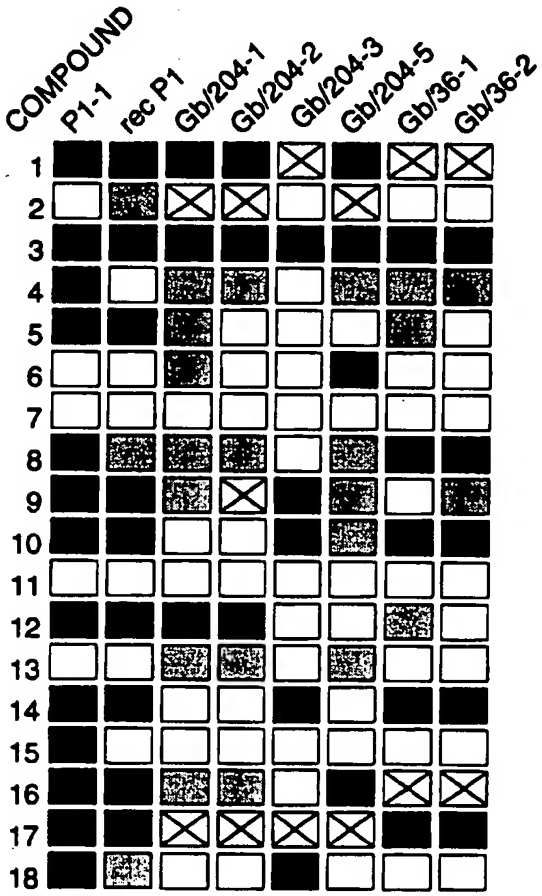


FIG. 3

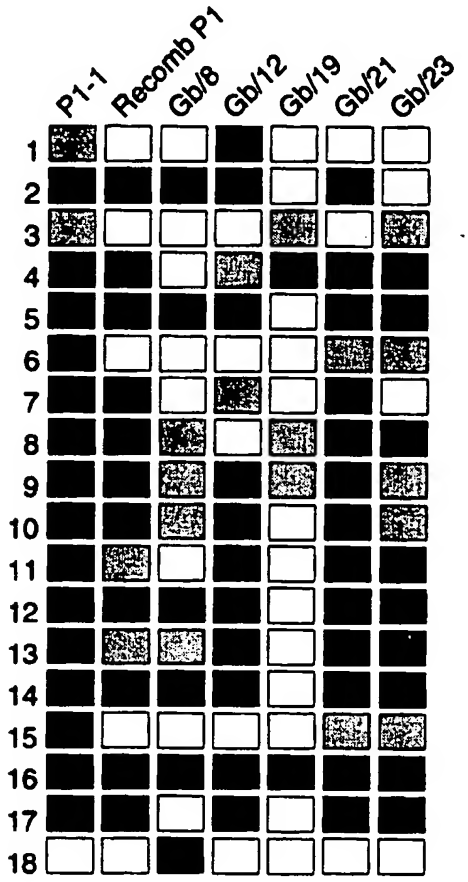


FIG. 4

FIG. 5

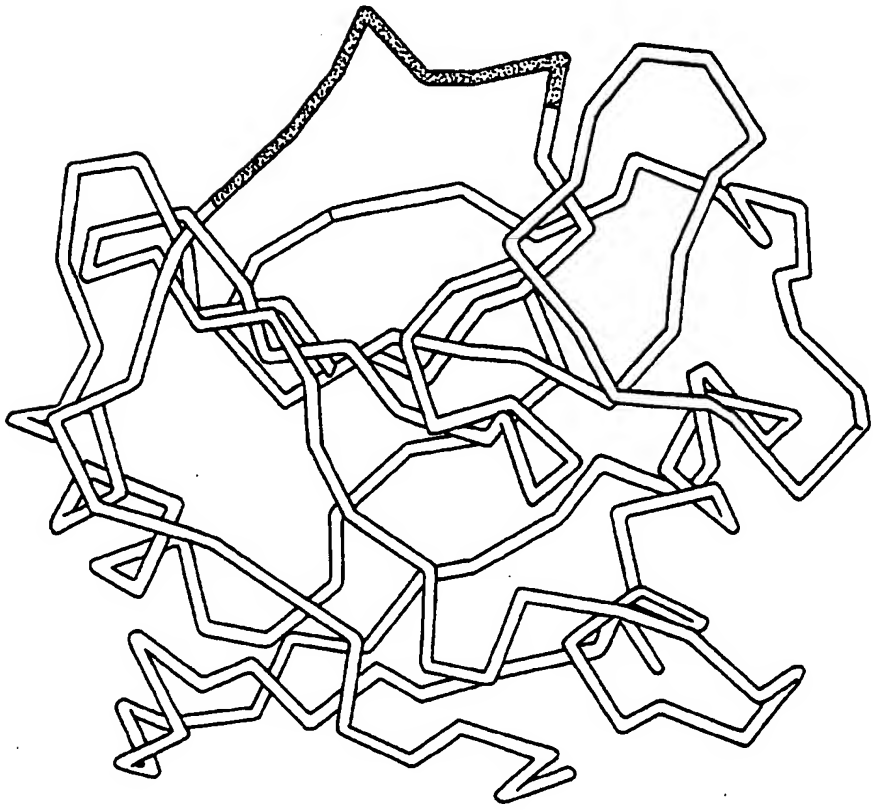


FIG. 6

INTERNATIONAL SEARCH REPORT

national Application No
PCT/US 96/01567

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/11 C07K14/245 C07K14/36 C07K14/42 C07K14/79
C12N9/06 C12N9/10 C12N9/90 G01N33/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BIOCHEMISTRY, vol. 29, no. 34, 28 August 1990, EASTON, PA US, pages 8063-8069, XP002005609 S HUANG ET AL.: "Role of lysine-54 indetermining cofactor specificity and binding in human dihydrofolate reductase" see the whole document ---	11
X	PROTEIN ENGINEERING, vol. 5, no. 7, July 1992, ENGLAND GB, pages 637-645, XP002005610 T DANDEKAR AND P ARGOS: "potential of genetic algorithms in protein folding and protein engineering simulations" see the whole document --- -/--	1

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
13 June 1996	10.07.96
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Authorized officer Masturzo, P

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO,A,90 07862 (PROTEIN ENGINEERING CO.) 26 July 1990 see the whole document	1
X	JOURNAL OF BIOCHEMISTRY, vol. 116, no. 1, July 1994, TOKYO JP, pages 34-41, XP002005611 K GEKKO ET AL. : "Point mutations of glycine-121 of E. coli dihydrofolate reductase; important roles of a flexibel loop in the stability and function" see the whole document	11
X	CHEMICAL ABSTRACTS, vol. 121, no. 3, 18 July 1994 Columbus, Ohio, US; abstract no. 29559s, A SIVAPRASADARAO & J FINDLAY: "Structure-function studies on human retinol-binding protein using site-directed mutagenesis" page 407; XP002005612 see abstract & BIOCHEM. J., vol. 300, no. 2, 1994, pages 437-442,	11
X	CHEMICAL ABSTRACTS, vol. 117, no. 11, 14 September 1992 Columbus, Ohio, US; abstract no. 106635k, B MASEPOHL ET AL.: "Functional analysis of the cysteine motifs in the ferredoxin-like protein FdxN of R. meliloti involved in symbiotic nitrogen fixation " page 358; XP002005613 see abstract & MOL. GEN. GEN, vol. 233, no. 1-2, 1992, pages 33-41,	11
A	WO,A,94 23298 (TERRAPIN) 13 October 1994 cited in the application see the whole document	20

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

ernational Application No
PCT/US 96/01567

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9007862	26-07-90	US-A-	5096815	17-03-92
		AU-B-	4958890	13-08-90
		CA-A-	2045516	07-07-90
		EP-A-	0452413	23-10-91
		JP-T-	4504052	23-07-92
		US-A-	5198346	30-03-93
WO-A-9423298	13-10-94	US-A-	5356784	18-10-94
		AU-B-	6622994	24-10-94
		CA-A-	2157929	13-10-94
		EP-A-	0692096	17-01-96